DOCUMENT RESUME

ED 177 536 CS 005 100

AUTHOR Hunter, William J.; And Others
TITLE "Which" Comprehension? Artifacts in the Measurement of Reading Comprehension.
PUB DATE Sep 79
NOTE 15p.; Paper presented at the Annual Meeting of the American Psychological Association (87th, New York, NY, September 1-5, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Content Reading; Differences; Elementary Education; *Evaluation Methods; Grade 5; Measurement Instruments; *Reading Comprehension; *Reading Diagnosis; *Reading Research; *Research Problems; Testing; *Testing Problems

ABSTRACT

Differences in 164 fifth grade students' reading comprehension scores were obtained with four different tasks typically used to measure comprehension (multiple choice, recall, cloze, and maze), using four different reading passages that were equated according to readability formulas. Data analyses revealed significant effects for passage, task, and an interaction between task and passage. It was concluded that the choice of a particular comprehension passage and testing procedure, whether in research or practice, may lead to conclusions that inappropriately generalize to other operational definitions of reading comprehension. These results suggest serious limitations of most contemporary reading comprehension research and testing. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

ED177536

Which Comprehension? Artifacts in the Measurement of Reading Comprehension

William J. Hunter
Mount Saint Vincent University

Janet Kendall
Simon Fraser University

Jana Mason
University of Illinois at
Urbana Champaign

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
William J. Hunter
Janet Kendall
Jana Mason
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS005100

Paper read at the annual convention of the American Psychological Association, New York City, September, 1979.

Which Comprehension? Artifacts in the Measurement of Reading Comprehension

Numerous tasks have been used to measure a reader's comprehension. For many years the multiple choice format has been widely used in standardized achievement tests and in basal reading materials. Alternatives to this measure include a recall task, in which, after people read a passage and are given an unrelated task such as mental arithmetic, they are urged to describe everything they can remember about the passage. Another alternative, the cloze test, is constructed either by deleting every fifth or seventh word or occasionally by deleting only content words. Cloze tests are usually scored by counting as correct only replacements of the exact words that were deleted. Maze tests are variants of the cloze procedure in which every fifth or seventh word is replaced by three choices, one of which is the correct word.

Regardless of the measurement task chosen, the researcher interested in comprehension must base an assessment of comprehension on memory of what was written and on completion of some task which is considered to have face validity. Often, researchers presume similarity of passages within a given study or comparability of results across different measures of comprehension. To do so requires two assumptions: equivalence of differing operational definitions of comprehension and equivalence of passages.

These assumptions of equivalence are not universally accepted. For example, Clark (1973) has argued that researchers must be cautious

in generalizing from results based on one set of materials to others - committing the language-as-fixed-effect fallacy. Moreover, Gibson and Levin (1976) state, ". . . almost any measure raises questions of validity, reliability, and generalizability to any other material or task (p. 409)." Even Golinkoff's (1976) caution regarding the assessment of comprehension skills in a number of studies of "good and poor" comprehenders, while commendable, might not be an adequate restriction on the interpretation of the results. In fact, it seems more likely that different task/passage combinations are not equivalent than that they provide similar measures of comprehension.

Different comprehension tasks, because they rely on memory search or organization of information, actually tap quite different strategies and processing mechanisms. Further, because passages are framed by differing structures and describe more or less familiar concepts there will be interactions between a reader's general knowledge and the passage content. Therefore, it could be true for any investigation of reading comprehension that different conclusions might have been drawn if other tasks and/or passages were used. Moreover, there could be even greater problems associated with comparing results from comprehension studies of populations that are deficient in skills that may impinge upon comprehensibility, such as decoding, since comprehension scores may then be hampered differently depending on the reading comprehension task.

The primary purpose of this study, therefore, was to examine the variations in reading comprehension scores that result from different

task/passage measurement combinations in order to evaluate the generalizability of results obtained in comprehension research.

Method

Subjects. Subjects were 171 fifth graders from four elementary schools located in a large eastern Canadian city. In order to assure comparability of participating schools, schools were chosen only if the average of the raw scores for fifth graders on the Metropolitan Achievement Test (MAT) Comprehension subtest corresponded to grade equivalents between 5.2 and 5.7. Seven subjects were randomly deleted in order to equate the groups. This resulted in 164 subjects, 41 per group.

Instruments. Multiple choice, recall, cloze and maze tasks were constructed for four reading passages from the McCall-Crabbs Standard Test Lessons in Reading (1961, first published in 1925). This set of reading passages was used since it has been the criterion for readability formulas more often "than any other single criterion" (Klare, 1963, p. 53). Also, there were many passages to choose from, all short, complete stories or expositions. Each passage was accompanied by 8 to 12 multiple choice questions which had been normed. We chose passages that, according to these norms were comprehensible at the 5.2 to 5.6 grade level when 70% of the questions were answered correctly. In order to make passages as equal as possible in other ways, we chose passages of 127 to 142 words, whose location of main idea was the first sentence and had a readability at grade 5 (Fry Readability Formula). However, we deliberately varied the text structure of the passages. One,

Woodpecker, is a fairy tale form. Ants is descriptive, though also in a story form. The other two are of an expository nature; one, Paper contains a list-like ordering of the steps in making of paper; the other, Whales is a causal structure, describing why people tagged whales.

Procedure. The passages and tasks were organized into a 4 x 4 Latin square design with students randomly assigned to one of the four groups. After receiving instructions and completing one example for each task, each student read the four passages, one under each of these four comprehension conditions: (a) McCall-Crabb multiple choice comprehension questions, with 9 or 10 items; (b) free recall, with students being asked to write down everything they could remember about the passage, scored as idea units with the highest score between 15 and 20; (c) cloze, with every fifth word being replaced by a blank and a maximum possible score from 24 to 28; and (d) maze with the same cloze word being replaced by three alternatives, one of which was the original text word, another a syntactically acceptable alternative, and the third a word from another part of the passage which was the wrong part of speech.

The students worked at their own pace with no time limits; the only constraint was that for the multiple choice and recall tasks they were not allowed to look back at the text. The number of correct responses was the dependent measure for three of the tasks. On the recall task, the number of idea units was measured.

Results

Data were analyzed following the procedures outlined in Winer's (1974) Plan 5. It is important to note that this design has a two-part interaction; one is a between-subject component and relates to group assignment, while the other is a within-subject component and relates to the fact that each subject provided data on all levels of each independent variable (task and passage). Winer regards this latter interaction as the more powerful of the two and suggests that in most cases it alone should be tested. Therefore, although both interactions are reported, only this within-subject component will be discussed.

In order to obtain a common metric for analysis, all scores were converted to proportions correct.[2] The proportions correct obtained by each student were averaged within groups with each group representing a unique task/passage combination. The averages are reported in Table 1. Inspection of these data shows that the maze and multiple choice tasks were consistently easier than recall and cloze. This observation is supported by the analysis of variance (Table 2); the difference among tasks was highly significant, $F(3,480) = 835.68$, $p < .001$. This was expected but was not a great interest (see discussion).

Of more interest were the passage and interaction effects. Referring again to Table 1, the story-form passage Woodpecker, was consistently easier than the others for all task types. Ants, which is nearly in a story form, was always second or third in difficulty, and Whales and Paper, the two expository passages, were the most dif-

ficult. Once again, the analysis of variance supported these observations, $F(3,480) = 17.80$, $p < .001$, showing that proportions correct differed significantly for the four passages. The third result an interaction between task and passage, $F(6,480) = 4.11$, $p < .001$, implies that the difficulty of the Whales and Paper passages depends on the type of measurement task employed. Paper was the more difficult to comprehend with the cloze task while Whales was the more difficult passage for the recall and maze tasks.

To examine the practical implications of these findings, students socring one S.D. or more below the mean were identified for each task/passage combination, yielding sixteen lists of "students in need of remediation" for each of the four groups studied. To determine the extent of agreement between any two task/passage combinations, the percentage of students on any given list whose names also appeared on the other list was computed. These percentages ranged from 0% (for eight of the 64 combinations) to 80% (a list of ten students scoring low on multiple choice/Paper included the names of four students out of five listed as scoring low on Cloze/Whales). The average of all combinations was 35%, showing only modest agreement in the identification of students who might need remedial work.

Insert Tables 1 and 2 about here

Discussion.

The finding that the maze and multiple choice tasks were consistently easier than recall and cloze is not surprising (Bormuth, 1968;

Rankin and Culhane, 1969). It is partly a function of the contrast between "selecting" from given answers (maze and multiple choice) versus "supplying" an answer (recall and cloze). Thus, the statistically significant difference among tasks is of little practical importance.

The finding that performance varied significantly on the four passages is of practical importance. Frequently both researchers and practitioners regard results based on different reading passages to be comparable. The present study suggests that this is not always the case, even when passages are as carefully selected and equated as those used here. The comparability of various studies in which similar tasks but different passages are used is questionable. Further, the validity of a single study in which different reading passages are used should also be questioned.

It is the interaction between tasks and passage though which is the most damaging to the notion of text/task compatibility. The combinations of the various tasks and passages produced systematically different results. This implies that each combination of reading passage and comprehension task may constitute a different operational definition of reading comprehension which could lead to different conclusions about comprehensibility or subjects' levels of comprehension depending on the passage/task combination involved. Thus the problems associated with comparing studies using different operational definitions may be even greater than previously suspected. It is conceivable that expository passages have differential effects depending on their cohesiveness. Paper had well sequenced information while Whales presented

the conclusion first and motivation information last. These kinds of deviations have been shown to affect recall (Stein and Negwaski, 1978) and a story reconstruction task (McClure, Mason, and Barnitz, 1979). Such text/task interactions, therefore, seem to warrant further consideration.

One question raised by the interaction between passage and task is which, if any, of the operational definitions used here is actually "reading comprehension." The answer, of course, is that each of them is, and, at the same time, none of them is. That is, as Davis (1968) has argued, comprehension is not a unitary construct. Several implications of this conclusion are warranted. First, researchers should be cautious about presenting research which purports to label different passages as equal; critical analysis of materials is necessary as well as a willingness of researchers to make their materials readily available to others. Second, further research is needed to determine what other factors, such as text cohesiveness, text structure, or topic familiarity, affect passage difficulty and interact with measurement of reading comprehension. Third, teachers ought to consider a variety of tasks and passage types when instructing students in reading comprehension since any one type of measure may be too restrictive. Fourth, these results suggest a need for more varied comprehension measures for testing and evaluating children. It is apparent that the difficulty found here in obtaining comparable tasks and measures makes comprehension test interpretation risky.

References

Bartlett, F.C. Remembering. Cambridge, England: Cambridge University Press, 1932.

Bormuth, J.R. Comparable cloze and multiple choice comprehension test scores. Journal of Reading, 1967, 10, 291-299.

Bormuth, J.R. The cloze readability procedure. Elementary English, 1968, 45, 429-436.

Bransford, J.B., & Johnson, M.K. Contextual prerequisites for understanding. Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 717-726.

Clark, H.H. The language-as-fixed effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 335-359.

Davis, F.B. Research in comprehension in reading. Reading Research Quarterly, 1968, 4, 499-545.

Gibson, E.J. and Levin, H. The psychology of reading. Cambridge: Mass.: M.I.T. Press, 1975.

Golinkoff, R.M. A comparison of reading comprehension processes in good and poor comprehenders. Reading Research Quarterly, 1975-76, 11, 623-659.

Guthrie, J.T. Reading comprehension and syntactic responses in good and poor readers. Journal of Educational Psychology, 1973, 65, 294-299.

Klare, G.R. The measurement of readability. Ames, Iowa: Iowa State University Press, 1963.

McCall, W.A., & Crabbs, L.M. Standard test lessons in reading. New York: Teachers College Press, Columbia University, 1961.

McClure, E., Mason, J., & Barnitz, J. Story structure and age effects on children's ability to sequence stories (Tech. Rep. No. 122). Urbana: University of Illinois, Center for the Study of Reading, May, 1979.

Rankin, E.F., & Culhane, J.W. Comparable close and multiple choice comprehension test scores. Journal of Reading, 13, 193-198.

Taylor, W.L. Cloze procedure: A new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.

Winer, B.J. Statistical principles in experimental design. New York: McGraw-Hill Book Co., 1971.

Footnotes

[1]Materials are available from the first author on request: Faculty of Education, Simon Fraser University, Burnaby, B.C.

[2]An arc-sign transformation was not applied to the proportions since analysis of variance is robust for departures from normality.

Table 1

Average Proportions Correct for Each

Task/Passage Condition

Passages	Cloze	Multiple Choice	Recall	Maze	
	Tasks				
Whales	.465	.856	.276	.862	.615
Woodpecker	.554	.951	.378	.938	.705
Ants	.510	.835	.334	.935	.654
Paper	.426	.816	.374	.906	.631
	.489	.864	.341	.910	

Table 2

Results of Analysis of Variance

Source	ss	df	MS	F
Between-Subjects				
Groups	.07	3	.02	.435
Subjects within groups	7.35	160	.046	
Within-Subjects				
Passages	.79	3	.263	17.08
Tasks	38.65	3	12.88	835.68
Passages X Tasks	.38	6	.063	4.11
Error	7.4	480	.0154	